

# **DATA COLLECTION METHODS IN THE STPDS SURVEYS: IMPROVING RESPONSE RATES IN MIXED MODE SURVEYS**

**R. Keith Wilkinson & Catherine Jucius Hines**

## **INTRODUCTION**

By the end of the 1980s all three of the personnel surveys of scientists and engineers (S&E) that make up the Scientific and Personnel Data System (STPDS) were experiencing severe response problems. By 1986 or 87 the unadjusted response rates were between 44 and 58 percent of the initial samples, rates that clearly indicated unacceptable levels of potential response error. Between 1989 and 1992 NSF began incorporating methodological tests into survey waves or pretests. There are enough similarities in the three surveys that it has been possible to make generalizations about the impact of methodological changes across the 3 surveys based on a test during a single panel or pretest. This has allowed NSF to test a large number of data collection improvements in just a few years, and select the most promising ones for the series of interrelated surveys that will begin in 1993.

The three surveys (the National Survey of College Graduates (NSCG), the Survey of Doctorate Recipients (SDR) and the Survey of Recent Science and Engineering Graduates (New Entrants)) each focus on a slightly different subpopulation of the nation's scientists and engineers. Each survey uses a slightly different sample frame, to cover slightly different components of the S&E population. The NSCG is based on a decennial census frame that includes reported occupation, to reach the 2% of the general population with S&E occupations more efficiently. The SDR employs a list of Ph.D.s from the National Research Council's Doctorate Record File to reach new doctorates, and maintains a longitudinal subsample back to 1941. The New Entrants survey uses administrative records from colleges and universities to reach recent recipients of bachelor's or masters degrees. All three surveys use primarily self-administered mail questionnaires, and had comparable response problems related to contact, non-response, and response bias.

## **MAIN DATA COLLECTION PROBLEMS IN THE 1980s**

The main shared problems fell into 3 areas: Locating respondents, motivating response, and maintaining longitudinal response. The first problem, locating respondents, was most significant for each survey during any respondent's first year in sample.

Motivating response, especially differential motivation between S&E and nonS&E respondents, caused significant error in all three surveys throughout the decade. For the NSCG, which included in the 1982 survey a substantial number of sample cases known to have a low probability of being outside S&E by training or occupation, lower response rates among non S&E had a substantial impact on overall response rates as well as response bias. For the other two surveys, which focussed on persons known to have degrees in S&E, the impact was more subtle. Non-response studies indicated that persons who were not currently employed in S&E occupations were more likely to be non-respondents, causing bias in estimates of S&E career paths.

Finally, errors related to sample attrition became increasingly important over time in the two longitudinal surveys (the NSCG and SDR). This had a more obvious affect on the later waves of the NSCG, due to the decision not to attempt to contact respondents who failed to respond to a previous wave. But it also had an impact on longitudinal response rates in the SDR, where the oldest cohorts had been asked to respond over four decades.

## **IMPROVING RESPONSE RATES IN THE 1980s**

The CNSTAT analysis of response error in the STPDS surveys began with a look at the response rates that had been achieved in similar studies (national longitudinal samples of persons with higher education) during the 1980s. They concluded that with improvements in data collection methodology, NSF could expect response rates high enough to ensure acceptable data quality over the decade of the 1990s.

CNSTAT stated that surveys achieving higher rates utilized features such as better content or appearance of contact materials or incentives to improve motivation, multiple mailings, and intensified phone follow-up of mail non-respondents. They recommended use of Dillman's Total Design Method to improve the appearance and schedule of mail packages, and concluded that higher response rates for the NSF surveys could be achieved, although with commensurate increases in per case costs.

The CNSTAT recommendations were received in time to begin incorporating tests of the cost and effectiveness of improving data collection methods in the SDR survey in 1991, and in a pretest of the NSCG conducted by the census in 1992. In general, each improvement that resulted in improved response rates or reduced response bias was incorporated in the survey panels that followed. The first steps involved non-response bias studies in the 1989 SDR and New Entrants surveys, then experiments with personalization and non-monetary incentives in mailing packages, and finally tests of instrument length, TDM mailing packages and monetary incentives in the 1992 NSCG pretest.

### A. To reduce problems with locating respondents

1. NCOA: In 1989, CNSTAT recommended improving sample contact rates through use of postal checks for address updates before the survey questionnaires were mailed. This procedure takes a minimum of several months, and the quality varies geographically. Instead, the census has recommended using a new update procedure that was not available in 1982, the use of an automated vendor file of current postal change-of-address updates, the National Change of Address (NCOA) file. These files are updated as often as every 2 weeks, and maintain a record of previous address updates (and, in some cases, telephone numbers).

The 1992 NSCG pretest was designed to use a sample frame with addresses about 3 years old at the time of mail contact, to evaluate the impact of an automated NCOA update of 1990 addresses in 1993.

If NCOA proves successful in the NSCG pretest, automated updates of sample files using current vendor files should be standard for each wave of the STPDS surveys. NCOA updates (for both address and telephone number) are far faster and less costly than multiple mailings or telephone contacts, so NSF also anticipates using them to maintain up to date address and telephone lists between longitudinal survey waves.

The 1992 NSCG pretest will be one of the first tests of this approach at the census, so there is no record of impact, but the bureau anticipates a significant reduction of non-response through updating out-of-date address lists with an NCOA vendor. A small scale study of NCOA updates for the highly mobile New Entrants population is also underway.

Preliminary results from the initial NCOA update of addresses in the NSCG pretest indicate that about 10 percent of sample addresses were updated before mail out. A tally of undeliverable addresses will be available soon, and permit us to estimate the impact of the NCOA update on locate rates and mail-response rates. The extremely good overall mail response rates achieved to date in the pretest indicate that the NCOA may have had a significant impact on non-locate problems.

2. Look up operations: For the NSCG in 1982 almost half of the non-mail-response cases referred to telephone follow-up were ultimately non-responses, because no telephone number could ever be located. The SDR experienced similar non-contact rates for telephone follow-ups in 1989 and 1991. They had some success supplementing directory assistance with a proprietary service that provides neighbor's names and phone numbers for sample addresses. This service may be tested during telephone follow-up in the NSCG pretest. The Census bureau is currently updating the file of non-mail-respondents to the pretest with an NCOA vendor who provides telephone numbers as well as address updates. The effectiveness of this update will be evaluated after telephone follow-up (scheduled to begin August 15), when the census will provide tallies of the source of phone numbers that result in telephone contacts; i.e., NCOA phone number, directory assistance, or other procedures.

3. Personal Visit follow-up: Personal visit follow-ups were not used in any of the surveys in the 1980s, except in one small-scale study of non-response bias (in the 1982 NSCG). One of the options that NSF is considering for the 1990s is personal visit follow-up to maintain extremely high response rates, or to measure or control response bias. Although these methods increase data collection costs dramatically (census has estimated that personal visit follow-ups could account for up to 1/3 of all data collection costs, even for a tiny minority of cases, in 1993), they will be tested with mail and telephone non-respondents in the pretest, to measure their impact on reduction of non-response or non-response bias. Personal visit interviews are not scheduled before October, so the results are not available at this time.

#### B. To reduce non-response

While the documented refusal/nonresponse rate for respondents in any of the surveys who were known to be contacted was low (6 percent in the NSCG in 1982, under 4 percent in the 1987 New Entrants), every possible measure that might improve overall response rate has been incorporated in the 1992 NSCG pretest to measure the overall improvement in response rate that we can hope to achieve. NSF was extremely lucky that during the period for re-design of the STPDS surveys Don Dillman has been at the Census Bureau in the capacity as a senior advisor and that he was available to serve as a member of the data collection advisory group to the STPDS even earlier.

1. Personalization: Contact materials for each of the surveys have been prepared using personalization since a 1989 test during SDR follow-up. In that test respondents receiving personalized mailing materials responded at a rate almost 8 percent higher than others, and participants in focus groups for each of the surveys have been equally as enthusiastic in preferring personalized mailing materials. For the NSCG pretest, materials are being prepared with the name of the sample person, drawn from the census files. Letters are varied to reflect the sequence of events and the respondents' response status. The letters are signed by the census personnel responsible for the survey.

2. Compressed mailing schedule: The NSCG pretest mailing schedule was designed to accomplish the sequence of mailings as quickly as possible, given the time needed to update files to reflect checked-in materials. The blanket prenotification letter was mailed less than a week before the first mailing package, and the blanket thank you/reminder card was mailed about 5 days later. To decrease the chance that the USPS or the respondents would set the mailings aside, all the mailings went with first class postage. Only 28 days were allowed between the first and second questionnaire mailings, and a further 28 days was scheduled before the mail-non-respondents are referred to telephone follow-up. According to the TDM model, the compressed schedule reduces the probability that a mailing package will be forgotten or discarded before it is answered.

3. Incentives: In the 1991 SDR the impact of a non-monetary incentive on mail-response rates was tested. Three treatments were used, one in which an informational brochure reporting previous survey results was sent with a pre-notification letter, one in which the same brochure was sent with the first questionnaire mailing, and one in which no brochure was sent. The results of this test were hard to interpret, because the group that received the brochure in the pre-mailing performed slightly better than the group that received no brochure, and the group that received the brochure with the questionnaire had the lowest return rate, by a slight percentage.

A second test using a monetary incentive was built into the 1992 NSCG pretest. Unlike the other surveys, which bring new graduates into their samples over time, the 1993 NSCG is the only source of longitudinal samples throughout the decade. Without extremely high response rates in the baseline survey, it is difficult to maintain strong response rates in the later panels. For this reason the 1992 pretest included a test panel that received a \$5 check enclosed with the first questionnaire mailing package. The pretest was designed to measure the impact the incentive had on mail-non-response as well as final response and on quality and completeness of response. Given the costs of incentives (in dollars and in public perceptions), the pretest was designed to indicate whether the impact of the incentive was sufficient to outweigh the costs, and whether its performance was sufficiently superior to outweigh the benefits associated with using another treatment.

Preliminary pretest results indicate that the incentive had a significant impact on early mail return rates, but that the other treatment groups have tended to reduce that difference over time. Any measure of the impact on final return rates or data quality will not be available for some time, but the results were sufficiently different that this treatment group was included in the sample for the Census Bureau to follow-up in the current pretest.

4. Mailing package appearance: A great deal of effort, on the part of NSF, survey contractors, and expert advisors, went into re-design of the appearance of the mailing packages for the 1993 surveys. Some of the features, such as personalization, have now been standard since 1991. Design features for the NSCG pretest questionnaire were the result of new desktop publishing software and printing resources available to the census. The results were creative use of spacing, lettering and color to make the questionnaire as easy as possible for respondents to follow, and reducing to a minimum features introduced solely for the convenience of processing.

Some of these modifications have been most difficult for the Census to implement, because the scale of their operations makes it much more expensive for them to use new procedures that require different equipment or staff for processing. But for the NSCG pretest the census has used the highest quality stationery and envelopes and postage resources it had available, to give the mailing packages the look of professional correspondence. The results are more like commercial professional mailings than the standard census package, and may have contributed to the overall high response rates in the pretest.

5. Questionnaire length: The goal of the STPDS surveys is to use similar instruments or items as much as possible, to make composite estimates more useful. By 1992 a baseline STPDS questionnaire was prepared in time for the NSCG pretest. Participants in early focus groups using the proposed 16 page questionnaire frequently objected to the time required to complete the long questionnaire. Although previous census research had indicated that length of questionnaire alone did not reduce response rates (KAP, 1984), the concern for overall response rates for 1993 caused NSF to test the impact of questionnaire length on response rates. The NSCG pretest incorporates questionnaires of 3 lengths, long, mid-length and screener, which differ largely by the incremental addition of items. The three lengths (4 pages, 12 pages and 16 pages) were chosen to enable NSF to project the cost (in diminishing response rate) of additional data items.

In addition, the screener version allows NSF to measure the impact of a two-stage data collection methodology, in which the brief screener is used to identify S&E respondents, who receive a follow-up questionnaire that incorporates the balance of the long form items. This approach is potentially an advantage only in the NSCG, which includes the non S&E college educated population in its sample. A screener collects detailed data only for the population of greatest interest (those with S&E training or employment), once the population size is established by a probability survey. The potential disadvantage is the attrition between the first and second questionnaires.

Preliminary NSCG pretest results indicate that there is an inverse relationship between questionnaire length and mail-non-response, but this may decrease after telephone follow-up. Preliminary results indicate that the screener mailing did achieve an excellent mail-back rate (second only to the incentive mail-back rate), and very preliminary results indicate that second-stage mail-back for this treatment is surprisingly high (64 percent after less than 2 weeks).

Overall, the response rates for the long-form are sufficiently good that it can be used as the basis for the 1993 surveys without anticipating any sacrifice of overall response rates.

6. Follow-up: The commitment of resources for telephone follow-up and (potentially) for personal visit follow-up should ensure improvement to overall response rates for all the surveys. The impact in the 1991 SDR survey has been positive. After 17 weeks of telephone follow-up the response rate is 79.6 percent based on complete CATI telephone follow-up. The impact of complete CATI telephone follow-up in the NSCG pretest won't be known until after August. Currently, telephone follow-up is scheduled for 400 pretest cases drawn from all treatment groups except the mid-length form. Response rate improvements for that form length were minimal, and it is no longer being considered as a design alternative for 1993.

### C. To reduce SE/nonSE bias

It was not difficult to identify features in the design of the surveys in the 1980s that might have contributed to the differential response rate between S&E and nonS&E respondents in the samples. The biggest problem existed for the NSCG, which includes in the sample many persons who had neither S&E training nor occupations. The problems were more limited in the SDR and New Entrants, where it appears that content materials should be modified to make it clearer to persons in non-S&E occupations that the survey does apply to them, their response is important, and they should fill out and return the questionnaire. More extensive changes have been required for the NSCG.

1. Titles: The new title of the postcensal survey is the National Survey of College Graduates, a title selected to emphasize the characteristic that all respondents in the baseline survey share - college degrees - and to eliminate reference to science or engineering. In focus groups with non S&E persons, this title appeared to avoid the problems associated with the earlier title.

2. Modifications to contact materials: The original plans for the NSCG pretest and the 1993 survey, based on expert advisory consultations and focus group results, was for the contact materials to come from the Census Bureau, and stress the Census Bureau as the vehicle, rather than the NSF as the sponsor. During clearance of the pretest, NSF was required to re-cast the pre-notification letter to stress NSF sponsorship, and to use NSF stationery and an NSF signature. Unfortunately this occurred at a point when it was too late to re-design the envelopes for the pre-notification letter, so the letter went out in a census bureau envelope.

It is difficult to know what impact the mis-match between stationery and envelope may have had on some respondents, or the impact of the emphasis on NSF in the first contact letter. We do know, however, that in each treatment group in the NSCG pretest S&E respondents out-number non S&E respondents by several percentage points. It is NSF's contention that this can be eliminated, and the S&E/nonS&E response bias reduced, if we return to the census bureau stationery in the 1993 survey.

3. Questionnaire content: Briefly, the appearance of all the questionnaires was modified in an effort to prevent major content items from reducing the probability of response from non S&E respondents, and to make it more likely that out-of-scope respondents (or their proxies) will complete and return the questionnaires.

#### D. To improve longitudinal response rates

For each of the surveys NSF plans to test the effectiveness of using mail contacts between survey waves to maintain up-to-date addresses. If these contacts include informational brochures that convey the importance and uses of the survey data, it may have some positive impact on respondent motivation. In addition, NSF will continue to test the effectiveness of NCOA address and telephone number updates on the control of sample attrition over time.

### **CURRENT STATUS AND FUTURE RESEARCH**

The response rates for each of the surveys since 1989 have improved substantially. In many cases the contribution of each individual improvement can't be measured, but telephone interviews of mail-non-respondents in the NSCG will include questions about reasons for non-response, which may help NSF understand which improvements have been the most important.

After CATI follow-up the 1991 SDR has achieved an 80 percent response rate. Follow-up to the 1992 NSCG pretest is not complete yet, so NSF cannot predict the final overall response rates. Most of the changes in data collection methods related to telephone or personal visit cannot be evaluated yet. But the results for the panels since 1991 indicate that the response rates in 1993 will be high enough to ensure good quality data from the individual survey waves and from the composite estimates throughout the 1990s.

Results to date clearly indicate that the major data collection improvements NSF has been testing should be incorporated in the plans for 1993. These include NCOA updates of the sample frame, as well as to ascertain telephone numbers and to update addresses between survey waves; improving the appearance of all mailing materials; removing sponsorship bias from contact materials and questionnaire appearance; including a pre-notification letter, compressed schedule, and reminder card in the mailing package; and investing in complete telephone follow-up for mail non-respondents, including development of a CATI questionnaire.

To increase the comparability of data from the three surveys in 1993, all are scheduled to be mailed out simultaneously. NSF will no longer be able to test methodological changes in sequence, applying the lessons of the most recent survey to those that follow. But the similarities across the surveys between the data collection procedures, the questionnaires, and the kinds of respondents, means that in many cases results of a test in one survey will still be generalizable to the others. This means that small scale tests of data collection methods can be built into any one of the surveys in NSF's efforts to improve the quality of the STPDS data.